

Web Based Management Information System With Optical Character Recognition Technology For A Phlippine Accounting Firm

Jinky B. Tumasis^a

^aTaguig City University, Philippines

Corresponding author.

Correspondence: Jinky B. Tumasis
Email: jtumasis2014@gmail.com

Article info

Received 2th December 2021 Receive
in revised form 27 February 2022
Accepted 18 March 2020

Keywords

OCR Technology, Management
Information, Breakthrough
Adaptation

Abstract

This paper describes the design of an Online Management Information System using a website connected to a portal with the adaptation of an OCR (Optical Character Recognition) Technology for Philippine Accounting Firm. The company often handles manual written and printed forms for preserving customer profiles and records of clients' registration required to various government agencies entities such as BIR, SSS, Philhealth and Pag-ibig Forms which were manually kept in the storage upon registration and if needed for data gathering or update, documents were retrieved in the storage manually by the employee since there is no software application developed yet to track the records easily for adding, updating and monitoring each client's details. This proposed system composed of two parts: a website and a portal where the website will be deployed in the World Wide Web with a unique domain and server which the clients can view and contact for the firm's services and a connected external link to access the portal's main system accessed by the firm's employees (upon registration and approval for signing in) only. The portal has two ways to manage the client's record using manual keying or adapting the OCR Technology using a scanner as an input device to scan the forms and convert directly to text input in the portal without inputting manually and individually for easy data storage and update. This system provides an accurate, complete monitoring and online management portal recording system with the adaptation of OCR Technology, eliminates the time spent for manual adding, updating and retrieval of client's records, and tracks the records easily using its enhanced system to easily add, update and monitor each client's details. This system will also serve as a draft or a starting point for the OCR Technology for application on accounting businesses and benefit other researchers who wish to have similar studies related to OCR Technology.

I. INTRODUCTION

Management Information System (MIS) is a computer system including both hardware and software which serves as the backbone of an organization's operations and processes. It gathers data from different types of systems (online or offline), analyzes the information, and reports data to assist in organization's decision-making (Shopify, 2019). In the field of accounting, this system is used to manage and track business' accounts records such as summary of clients' profiles, financial transactions including expenses, sales, liabilities, etc. and is capable of producing comprehensive statistical reports that provide management or interested parties with a clear set of supporting data to aid in the decision-making process due to complex series of manual calculations and balances. The most common types of accounting systems are manual and automated or computer-based systems. Manual accounting systems are time consuming and involve more paperwork activities. Since everything is done manually, there's a higher risk of human and process error.

Currently, most companies are using automated and computer-based systems using websites connected to a database to organize manual collection of data, stored and can be accessed electronically using different kinds of devices (e.g. (Computer systems, laptops, and mobile phones). Database designers typically organize the data to model aspects of reality that supports processes requiring information using specialized

software deployed locally and/or cloud-based services. With the continuous improvement of technology, one of the famous concepts introduced of the image processing, is the Optical Character Recognition (OCR), defined as the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text from a scanned, photo or any image-type text-type document. This technology is widely used as a form of information entry from printed paper data records – whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation – it is a common method of digitizing printed texts so that they can be electronically edited, searched, compressed, presented online, and used in machine processes such as cognitive computing, machine translation, (extracted) text-to-speech, key data, and text mining.

The current record system uses manual written and printed forms for storing of customer profiles and records such as BIR, SSS, Philhealth and Pag-ibig Forms. These hardcopy records were kept in the storage upon registration and if needed for data gathering or update, documents were retrieved in the storage manually by the ~~staffs~~ along with the other records. No software applications were still yet ~~to be developed~~ to manage these ~~records~~. These activities add and ~~giving~~ more pressure on each of the ~~staffs~~ than attend and focus more ~~on the~~ main manual accounting computation process activities.

As a proposal, this website connected to a portal with the adaptation of an OCR (Optical Character Recognition) Technology for Philippine Accounting Firm. The system is composed of two parts: a website and a portal. The website will be deployed in the World Wide Web with a unique domain and server which the clients can view and contact for the firm's services and a connected external link to access the portal's main system.

II. LITERATURE REVIEW

Management Information Systems (MIS)

Integrated physical warehouse and digital document management system

A document management facility ~~processes the volume~~ of physical documents for digital imaging and ~~utilizes~~ robotic apparatus to perform the various functions for processing the documents, which include document intake, storage, digitization, and/or removal. It can be accessed with a document management system. The aspects of the invention directed to document digitization section includes a document preparation station configured to prepares and separates a plurality of documents; a fastener management station configured to detect and remove fasteners from the plurality of documents; and a document scanning station configured to digitize individual pages of the plurality of documents

Conversion of different types of documents to a uniform and an editable or a searchable format

A device may obtain a set of documents with different document types for conversion to an editable or a searchable format such as pdf, Microsoft Word files, etc. The device identifies a portion of the set of documents that is to be converted to editable text. And generate a set of converted documents based on the conversion information and the zone relevant data. Each document in the set of converted documents may be of the document type.

Image-based Data Management Method and System

Different methods and systems are provided for storing, organizing, accessing, and communicating using image-based ~~files~~. An automated or computer-based system implemented method for online communication using business cards includes receiving a business card in an image format, performing an optical character recognition (OCR) conversion procedure to create an equivalent business card in text format, and recognizing keywords on the equivalent business card in text format, linking the keywords with the business card in image format and the corresponding equivalent business card in text format, attaching user-provided privacy settings of access permissions to the business card, attaching user-provided contents to the business card, and delivering information about the business card and its associated contents to one or more users online

Process performance evaluation for Enterprise data systems

A method using integrated software and algorithms for measuring, modeling, benchmarking and validating any Enterprise Content management system, forms processing data capture system or data entry system, ingest of special engineered test components such as a Digital Test Deck®, as well as data quality grading algorithms, use of cost models, validation of downstream business processes, and implementing statistical process control (Dawson, 2011).

The invention relates to Enterprise Content Management (ECM) systems and other forms (including checks) processing for data capture systems such as manual data entry from image or paper, recognition technologies (OCR, ICR, OMR, etc.). It also enables real-time measurement and modeling of system characteristics as part of the workflow of a current Enterprise Content Management (ECM) system or any other forms processing data capture system, conveniently measure system performance at any point in time for

Enterprise Service Bus or other workflow manager, it can also calculate a cost per form based on measured data (for example, the OCR engine's operation point), as well as reducing the expenses associated with the conventional inspection methods by deploying statistical process control.

Computer-implemented PDF document management

Portable Document Format (PDF) is a file format developed by Adobe Systems Incorporate that allows documents formatted in PDF to be accessed for online viewing, for example, on a monitor screen, in the same manner for example, in the same manner that hard copies of the documents would be viewed off-line. PDF documents, for example, can be viewed page by page and scrolled through online based on the demands of the users. Applications such as Adobe Acrobat Exchange can be used to view PDF documents. or in most cases, Adobe Acrobat Reader is necessary.. Using such applications in content-based document retrieval procedures, where large numbers of PDF documents may need to be retrieved, significant time and effort may be invested in assessing if retrieved PDF documents are duplicates for continuous viewing.

Optical Character Recognition History

Optical character recognition (OCR) of machine printed Latin script documents is ubiquitously claimed as a solved problem. However, error-free OCR of defective or noisy text remains a challenge for current OCR algorithms. Character recognition based on segmentation is used in the most recent techniques.. This is difficult since segmenting degraded text is difficult in and of itself.

Segmenting handwritten document pictures into text-lines and words is a critical operation in optical character recognition. OCR stands for optical character recognition systems. Identifying and detecting printed characters using imagery, which is now a common feature in scanners and mobile devices.

Document automated classification/declassification system (Schoen, 1997)

The document automated classification/declassification system is a system for automatically classifying or declassifying military, intelligence, government, or industrial documents. Classification or declassification guidelines are inputs to the system which describe the sensitive information, the document(s) to be handled, and most are in electronic format, For example output from word processor or another digital format). A software program generates a database based on classification criteria or rules, which is then stored in the computer system. The document(s) to be processed are scanned, and the database is used to identify categorized sections of the papers, using a second software program (driven by the rules for determining classification levels.

Government Records Automation Systems

BIR Forms - JuanTax as An Alternative to BIR Releases eBIRForms Package (JuanTax Mobile Application with OCR Technology)

JuanTax is a Philippine based cloud tax software which is designed to help businesses when it comes to filing all local taxes such as VAT (Value Added Tax), Percentage Tax, Withholding Tax and Income Tax Returns. The special features of JuanTax are automation (Not only does JuanTax minimizes our task of entering data, it also does auto calculation as well as populating of fields such as schedules, carryovers and amortizations), it's running in the cloud (Since JuanTax runs with the use of the internet and cloud technology, we'd be confident that our data is automatically backed up, and we can access it anytime, anywhere), flexibility (JuanTax comes in two versions: Xero integrated version, which lets us automatically integrate our JuanTax account with Xero accounting software, and Standalone, which allows us to use all the functionalities of JuanTax even if we're using different accounting applications.

Electronic BIR Forms (eBIRForms)

The Electronic Bureau of Internal Revenue Forms (eBIRForms) was developed primarily to provide taxpayers with an alternative mode of preparing and filing tax returns that is easier and more convenient. The use of eBIRForms by taxpayers will improve the BIR's tax return data capture and storage thereby enhancing efficiency and accuracy in the filing of tax returns. With the downloadable eBIRForms Software Package (also known as the Offline Package), taxpayers and Accredited Tax Agents (ATAs) will be able to fill up tax returns offline and submit it to the BIR through the online eBIRForms System ("eBIRForms", N.D.).

III. RESEARCH METHODOLOGY

1. RESEARCH DESIGN

This study employed both descriptive and developmental methods. Using the descriptive method, it is used to describe the level of acceptability and efficiency of the developed online information system using the evaluation results gathered to both the Phillipines Accounting Firm's employees and five IT experts and utilized the developmental method because its primary aim was to develop a web based management information system

A. Descriptive Method

This study utilized the descriptive research method. According to Bermudo, et al., (2010), this research design allows the researcher to describe systematically, factually, accurately, and objectively a situation, problem or phenomenon as it naturally occurs to answer the questions and describe the data and characteristics of the subject being studied. It involved the description, recording, analysis and interpretation of the requirements from the start and end of the development stage .

B. Developmental Method

In this study, the developmental research approach for instructional technology was used, in which the research was combined in the development process (Ritchey, 1997). The study was divided into several phases which may be iterative: design, development, implementation and evaluation.

Project Development

The Fourth Generation Technique (4GT), which was being utilized to rapidly create data frameworks and systems. This procedure made the utilization of various programming advancement apparatuses.

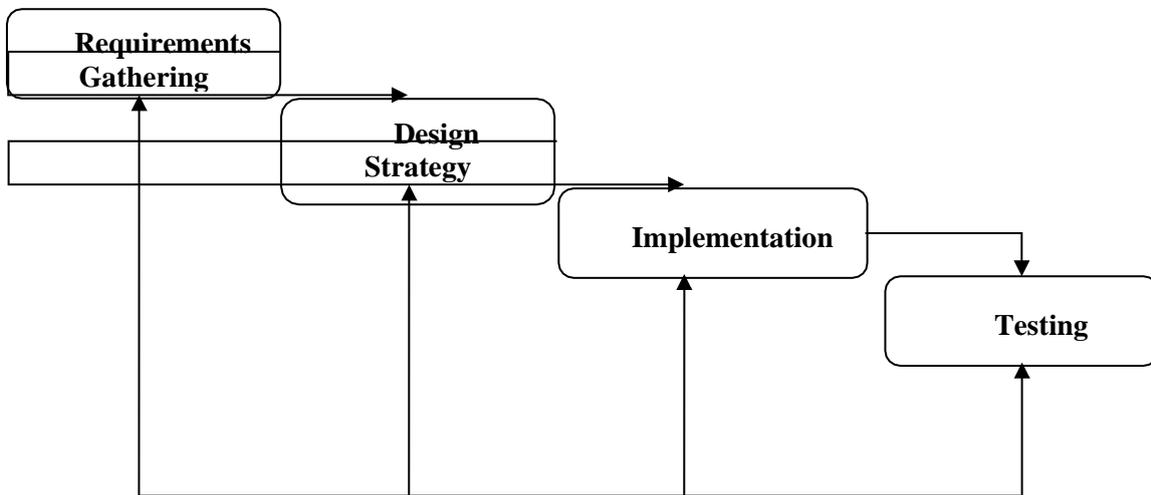


Figure 1. Fourth Generation Techniques (4GT)

Sources of Data

The study had two sources of data, namely; primary and secondary sources. The primary sources of data were the owner of the ~~Accounting~~ ~~Accounting~~ Firm and employees. With their cooperation during the interviews, they emphasized that they wanted a website and a portal to help them add, update and retrieve clients’ records easily. The secondary sources of data included the research through various documents related on the portal’s content together with OCR Technology through the studies of the researchers in the pasts and current projects published through various portals, journals, projects, etc.

IV. FINDINGS AND DISCUSSION

a. Testing Result

Using the overall testing procedures, the researcher together with the help of accounting firm employees and IT professionals involved in this study able to test the system into (2) categories:

2. **Overall system function and operational test** – Wherein the researcher tested the overall software quality of the system’s overall operation, functionality and confidentiality.
3. **Equipment / hardware test** – Wherein the researcher tested the performance of the system via different ~~platforms~~ ~~platform~~ and equipment.

i. Accounting Firms Employees’ Level of Acceptability of the System

Summary of System Acceptability:

Indicator	Weighted Mean	Verbal Interpretation
Functionality	3.20	Strongly Acceptable
Reliability	3.47	Strongly Acceptable

Usability	3.35	Strongly Acceptable
Efficiency	3.20	Strongly Acceptable
Maintainability	3.50	Strongly Acceptable
Portability	3.05	Strongly Acceptable
Overall Weighted Mean	3.30	Strongly Acceptable

ii. IT Professional’s Level of Effectiveness of the System

iii. Summary of System Effectiveness:

Indicator	Weighted Mean	Verbal Interpretation
Readability	3.80	Very High
Ease of program maintenance	3.13	Very High
Adaptation of the source code to other programming languages	3.00	Very High
Other systems considerations factors	3.13	Very High
Overall Weighted Mean	3.27	Very High

Findings

The overall findings of the study were as follows:

1. The current management information system of Accounting Firm uses manual written and printed forms for storing of customer profiles and records of client’s registration requirements to various government entities such as BIR, SSS, Philhealth and Pag-ibig Forms.
2. The online management system was an automated or computerized-based proposed system composed of *two parts: a website and a portal* where the website was deployed in the World Wide Web with a unique domain and server which the clients can view and contact for the firm’s services and a connected external link to access the portal’s main system accessed by the firm’s employees (upon registration and approval for signing in) only.
3. The firm employees’ evaluation results of the system’s acceptability showed the summary of the accounting firm employees’ evaluation of the developed online information management system.
 - a. The maintainability of the system ranked on top having the **highest mean of 3.5** considering the factors of user interface’ analyzability, changeability, stability and testability upon the usage of the accounting firm’s employees during evaluation run.
 - b. This was followed by the system’s reliability that has a mean of **3.47** considering the system’s fully functional upon implementation and deployment stage.
 - c. The system’s level of usability was on the third rank having a mean of **3.35** considering the high level of users’ understandability, learnability, operability and attractiveness of the system’s interface, functionality and efficiency of the system were on the fourth rank with a mean score of **3.2** and portability with different hardware and software systems with the mean of **3.05**.
 - d. The overall weighted mean obtained from all six criteria gained a mean score of **3.30** and standard deviation of **0.18** which was interpreted as **Strongly Acceptable** for both the accounting firm’s employees.
4. The IT Professional’s evaluation of the system’s effectiveness of the criteria of programs readability ranked at the top with a **mean of 3.8** followed by the ease of program maintenance and other considerations with the **mean mark of 3.13** and the least was the system’s adaptation of source code to other systems which has a **mean of 3.00**. Overall, the average mean of all criteria is **3.27** which was interpreted as **Strongly Acceptable**.

V. CONCLUSION AND FURTHER RESEARCH

In line with the evaluation of the developed system, the researcher found out the accompanying conclusions stated below:

1. The developer’s findings include the accounting firm’s employees and other people’s limited knowledge about the automated or computerized based system for management information system for storing, retrieving and updating records.

2. The developer developed an online management ~~system that aids~~ ~~system aids~~ the accounting firm that may help hasten the research and storage process of clients' profiles.
3. Based ~~on from~~ the evaluation summary of the accounting firm's employees, it was indicated that the developed system was strongly acceptable and found useful to improve the research and current manual recording system of the accounting firm.
4. The Online Management Information System was evaluated by selected IT Professionals and was found functional, reliable, usable, efficient, maintainable and portable to use.

REFERENCES

1. MANAGEMENT INFORMATION SYSTEMS (MIS)?. RETRIEVED FROM [HTTPS://WWW.SHOIFY.COM/ENCYCLOPEDIA/MANAGEMENT-INFORMATION-SYSTEMS-MIS](https://www.shoify.com/encyclopedia/management-information-systems-mis).
2. PICINCU, A . (2019) . *VARIOUS TYPES OF ACCOUNTING SYSTEMS*. RETRIEVED FROM [HTTPS://BIZFLUENT.COM/INFO-8213727-VARIOUS-TYPES-ACCOUNTING-SYSTEMS.HTML](https://bizfluent.com/info-8213727-various-types-accounting-systems.html).
3. ONDEMAND, HPE HAVEN . (2019) . *OPTICAL CHARACTER RECOGNITION* . RETRIEVED FROM [HTTPS://WWW.PROGRAMMABLEWEB.COM/API/HPE-HAVEN-ONDEMAND-OCR-](https://www.programmableweb.com/api/hpe-haven-ondemand-ocr-document)
4. DOCUMENT.
5. KANCHIRAYAPPA, V. N., & CHANDRAMOULI, N. (2005) . *COMPUTER-IMPLEMENTED PDF DOCUMENT MANAGEMENT* - U.S. PATENT No. 6,895,550 . WASHINGTON, DC: U.S. PATENT AND TRADEMARK OFFICE.
6. PATENT AND TRADEMARK OFFICE.
7. GHATAGE, P., DILLON, P., MISTRY, A., JONNAVITHULA, R., & THANGARAJ, N. K. (2018) .
8. *CONVERSION OF DOCUMENTS OF DIFFERENT TYPES TO A UNIFORM AND AN EDITABLE OR A SEARCHABLE FORMAT* - U.S. PATENT No. 9,886,436 . WASHINGTON, DC: U.S. PATENT AND TRADEMARK OFFICE.
9. HUANG, Y. C. (2012) . *IMAGE-BASED DATA MANAGEMENT METHOD AND SYSTEM* - U.S. \PATENT APPLICATION NO. 12/763,172.
10. DAWSON, J. W., JOHNSSON, E. T., & PAXTON, K. B. (2011) . *PROCESS PERFORMANCE EVALUATION FOR ENTERPRISE DATA SYSTEMS* - U.S. PATENT No. 8,055,104 .
11. WASHINGTON, DC: U.S. PATENT AND TRADEMARK OFFICE. SCHOEN, N. C. (1999) . *DOCUMENT AUTOMATED CLASSIFICATION/DECLASSIFICATION SYSTEM* -
12. *U.S. PATENT No. 5,991,709* . WASHINGTON, DC: U.S. PATENT AND TRADEMARK OFFICE.
13. PENG, G., YU, P., LI, H. AND HE, L. . (2016) . *TEXT LINE SEGMENTATION USING VITERBI ALGORITHM FOR THE PALM LEAF MANUSCRIPTS OF DAI* . 2016 INTERNATIONAL
14. CONFERENCE ON AUDIO, LANGUAGE AND IMAGE PROCESSING (ICALIP), SHANGHAI, PP. 336-340.
15. RYU, J., KOO, H. I. AND CHO, N. I. (2015) . *WORD SEGMENTATION METHOD FOR HANDWRITTEN DOCUMENTS BASED ON STRUCTURED LEARNING* . IEEE SIGNAL PROCESSING LETTERS, VOL. 22, NO. 8, PP.1161-1165, AUG.2015. DOI: 10.1109/LSP.2015.2389852. RASHID, S. F..
16. C. NARMATHA , DR. M. THANGAMANI , S. JAFAR ALI IBRAHIM, " RESEARCH SCENARIO OF MEDICAL DATA MINING USING FUZZY AND GRAPH THEORY", *INTERNATIONAL JOURNAL OF ADVANCED TRENDS IN COMPUTER SCIENCE AND ENGINEERING, VOL 9, No 1 (2020): 349-355*, [HTTPS://DOI.ORG/10.30534/IJATCSE/2020/52912020](https://doi.org/10.30534/ijatcse/2020/52912020)
17. SHAFAIT, F. AND BREUEL, T. M. (2012) . *SCANNING NEURAL NETWORK FOR TEXT LINE RECOGNITION*. 2012 10TH IAPR INTERNATIONAL WORKSHOP ON DOCUMENT ANALYSIS SYSTEMS, GOLD COST, QLD, 2012, PP.105-109. DOI: 10.1109/DAS.2012.77.
18. RAO, MR S. VENKATESWARA, MRS P. VIDYAVATHY, MR IMRAN MAHAMMAD, MRS M. RAJITHA KANTHA, S. JAFAR ALI IBRAHIM, AND NS KALYAN CHAKRAVARTHY. "YOUTH PSYCHOLOGY, THE IMPACT OF IDENTITY CRISIS IN SHAPING FUTURE IN CHETAN BHAGAT'S FIVE POINT SOMEONE." *TURKISH JOURNAL OF COMPUTER AND MATHEMATICS EDUCATION VOL 12, NO. 3 (2021): 3389-3394*.
19. AHMAD, I., WANG, X., LI, R. AND RASHEED, S. (2017) . *OFFLINE URDU NASTALEEQ OPTICAL CHARACTER RECOGNITION BASED ON STACKED DENOISING AUTOENCODER*. CHINA COMMUNICATIONS, VOL. 14, NO. 1, PP. 146-157.
20. CROFT, W.B., HARDING, S.M., TAGHVA, K., AND BORSACK, J. (1994) . *AN EVALUATION OF INFORMATION RETRIEVAL ACCURACY WITH SIMULATED OCR OUTPUT* . SYMPOSIUM OF DOCUMENT ANALYSIS AND INFORMATION RETRIEVAL, ISRI-
- UNLV.GOSHTASBY, A. AND EHRICH, R (1988) . *CONTEXTUAL WORD RECOGNITION USING PROBABILISTIC RELAXATION LABELING*. PATTERN RECOGNITION, 21(5):455-462.
21. PAL, P., KUNDU P.K., AND CHAUDHURI, B.B. (2000) . *OCR ERROR CORRECTION OF AN INFLECTIONAL INDIAN LANGUAGE USING MORPHOLOGICAL PARSING* . JOURNAL OF INFORMATION SCIENCE AND ENGINEERING, 16(6):903-922, NOVEMBER.
22. PEREZ-CORTES, J.C., AMENGUAL, J.C., ARLANDIS, J., AND LLOBET, R. (2000) . *STOCHASTIC ERROR - CORRECTING PARSING FOR OCR POST-PROCESSING*. ICPR, PAGES 4405-4408, BARCELONA, SPAIN, SEPTEMBER.
23. SINHAM, R.M.K. AND PRASADA, B. (1988) . *VISUAL TEXT RECOGNITION THROUGH CONTEXTUAL PROCESSING*. PATTERN RECOGNITION, 21(5):463-479.
24. TAGHVAM K. AND STOFESKY, E. (2001) . *OCRSPELL: AN INTERACTIVE SPELLING CORRECTION SYSTEM FOR OCR ERRORS IN TEXT* . IJDAR, 3(3):125-137.

25. TENELLEVEN (2017) . JUAN TAX AS AN ALTERNATIVE TO EBIR 7.1: A DEFINITIVE REVIEW . RETRIEVED FROM [HTTPS://WWW.TENELLEVEN.COM/BLOG/JUANTAX-ALTERNATIVE-EBIR-7-1-DEFINITIVE-REVIEW/](https://www.tenelleven.com/blog/juantax-alternative-ebir-7-1-definitive-review/)
26. R. GANESAN, M. THANGAMANI, S. JAFAR ALI IBRAHIM, “ RECENT RESEARCH TRENDS AND ADVANCEMENTS IN COMPUTATIONAL LINGUISTICS”, *INTERNATIONAL JOURNAL OF PSYCHOSOCIAL REHABILITATION VOL 24, NO 8 (2020):1154-1162*, DOI: 10.37200/IJPR/V24I8/PR280128
27. ELECTRONIC BIR FORMS (EBIRFORMS) . RETRIEVED FROM [HTTPS://WWW.BIR.GOV.PH/INDEX.PHP/ESERVICES/EBIRFORMS.HTML](https://www.bir.gov.ph/index.php/eservices/ebirforms.html)